Effectiveness of Structured Physical Activity Interventions Through the Evaluation of Physical Activity Levels, Adoption, Retention, Maintenance, and Adherence Rates: A Systematic Review and Meta-Analysis

Willinger, N., Steele, J., Atkinson, L., Liguori, G., Jimenez, A. & Horton, E.

Original citation & hyperlink:
Willinger, N, Steele, J, Atkinson, L, Liguori, G, Jimenez, A & Horton, E 2021, 'Effectiveness of Structured Physical Activity Interventions Through the Evaluation of Physical Activity Levels, Adoption, Retention, Maintenance, and Adherence Rates : A Systematic Review and Meta-Analysis', Journal of Physical Activity & Health, vol. 18, no. 1, pp. 116–129. https://dx.doi.org/ 10.1123/jpah.2019-0349

DOI 10.1123/jpah.2019-0349
ISSN 1543-3080

Publisher: Human Kinetics

Abstract

Background: Structured physical activity (PA) interventions (i.e. intentionally planned) can be implemented in a variety of facilities, and therefore can reach a large proportion of the population. Our aim was to summarise the effectiveness of structured interventions upon PA outcomes, in addition to proportions of individuals adopting and maintaining PA, and adherence and retention rates.

Methods: Systematic review with narrative synthesis and exploratory meta-analyses. Twelve studies were included.

Results: Effectiveness on PA levels during adoption (pre- to first time-point) showed a trivial standardised effect (0.15 [-0.06, 0.36]); during maintenance (any time-point after the first and >6 months since initiation) the standardised effect was also trivial with a wide interval estimate (0.19 [-0.68, 1.07]). Few studies reported adoption (k = 3) or maintenance rates (k = 2). Retention at follow-up did not differ between structured PA or controls (75.1% [65.0%, 83.0%] vs 75.4% [67.0%, 82.3%]), nor did intervention adherence (63.0% [55.6%, 69.6%] vs 77.8% [19.4%, 98.1%]).

Conclusion: Structured PA interventions lack evidence for effectiveness in improving PA levels. Further, though retention is often reported and is similar between interventions and controls, adoption, maintenance and adherence rates were rarely reported rendering difficulty in interpreting results of effectiveness of structured PA interventions.

Background

Physical activity (PA) levels in the UK are low, with 35.9% of the population currently being classified as insufficiently active [1]. In an international comparison these inactivity rates lie between those observed across Europe (29.4%) and North and South America (39.4%)[1]. Despite a multitude of interventions aiming to increase PA levels, these figures have not improved throughout past decades[1,2]. Many of PA interventions can be classified as structured interventions as they provide a clear recommendation on the frequency of attendance of pre-planned exercise sessions[3]. Structured programmes, if evidenced as effective in increasing PA, can potentially be implemented in a large variety of facilities in the private and public health sector, therefore contributing to the global target of reducing inactivity by 10% by 2025 as defined through the WHO[4]. However, from a public health perspective, though increased population average PA levels are no doubt desirable, it is the intention of organisations such as the WHO to increase the proportions of the population meeting PA recommendations. Changes in population average PA levels, even those of a large magnitude, may be driven by segments of the population. Thus, it is important to interpret the apparent effectiveness of interventions alongside the proportions of those meeting recommended levels of PA, and whom adhere to and are retained in interventions.

Adoption and maintenance of PA are frequently used terms, but their definition and measurement in academic literature vary widely[5–7]. Generally, adoption refers to an individual's uptake of PA at recommended levels, whereas maintenance is described as long-term behaviour change[5,6]. Due to the lack of explicit definitions of the term's 'adoption' and 'maintenance', in this review they were defined as meeting recommended PA levels; adoption referring to engaging in this behaviour upon, or shortly thereafter, the implementation of a structured intervention; and maintenance referring to the longer term continuation of this behaviour after adoption (for example, six months in accordance with previous literature defining this as necessary in order to establish long-term effects[8,9]). Similarly, retention and adherence or attendance, are often factors considered with respect to the implementation of PA interventions and likely influence their effectiveness. Retention typically refers to the number of participants available for follow-up assessments (independently of whether they have actually

maintained the intended PA behaviours or not), while adherence or attendance rates describe the degree to which participants take part in intervention sessions. Both retention and adherence are indicators for intervention implementation and feasibility.

Intervention effectiveness is commonly evaluated via the analysis of mean differences in outcome measures (i.e. PA levels) between treatment groups, and refers to the extent to which participation in an intervention can increase participant's PA levels. However, evaluation of intervention effectiveness, particularly with respect to implementation successfulness, should also consider the proportion of participants able to achieve and maintain the recommended activity levels, and whom are retained in the intervention and indeed adhere to it.

This review seeks to examine the effectiveness of structured PA interventions upon PA outcomes, in addition to proportions of individuals adopting and maintaining PA, and adherence and retention rates. The primary aim was to draw conclusions about the effectiveness of structured PA interventions and programmes in promoting the adoption and maintenance of PA in insufficiently active adult populations. In addition, we aim to draw conclusions on intervention implementation by summarising adherence and retention rates of participants undergoing structured PA interventions. The following research questions will be addressed: 1) What proportion of articles on structured PA interventions have evaluated and reported intervention effects for adoption and maintenance?; 2) What are the adoption and maintenance rates achieved by structured PA in insufficiently active adults?; 3) Are structured interventions effective in increasing PA levels compared to a control group or another treatment group?; and 4) What are the retention and adherence rates of participants of structured interventions and which conclusions on intervention implementation can therefore be made?.

Methods

The protocol for this review can be found under the PROSPERO registration number CRD42017061009, and therefore will only be outlined briefly in the following. We followed the PRISMA and AMSTAR 2 guidelines for the conduction and description of this review[10,11].

Literature Search Strategy

A combined search of five EBSCO databases (MEDLINE, SPORTDiscus, PsycINFO, CINAHL, Academic Search Complete) was carried out in addition to separate searches through Scopus and the Cochrane library from the earliest available date until May 2020. The search strategy can be viewed in full on PROSPERO. Terms related to 'adoption', 'intervention', 'physical activity', 'maintenance', 'retention' and 'adults' were combined to identify relevant articles. Terms related to 'nutrition', 'workplace', 'mass media' and 'children' were excluded from the search.

Reference lists of eligible articles were screened for relevant articles. Abstracts and full-texts of articles with relevant titles were screened and their eligibility was determined through a comparison to the inclusion criteria as detailed below. In addition, a separate search of grey literature in the form of evaluation reports of PA programmes in the UK was carried out.

Inclusion Criteria

The studies and reports had to meet the inclusion criteria detailed below. Only articles published in English language were included. In order to widen the evidence, a pragmatic approach was taken, and this review also includes study designs other than randomised controlled trials e.g. prospective cohort studies. This is taken into account when interpreting interventions results through the assessment of methodological quality of included articles through an evaluation of risk of bias according to Cochrane Guidelines, as detailed in the following paragraphs. Inclusion criteria were as follows: 1) The mean age of participants lay between 18-64 years; 2) Participants were characterised by an insufficiently active lifestyle at baseline, defined through activity levels of less than 150 minutes of moderate to vigorous intensity PA per week; 3) The population sampled was healthy adults independent of their weight status; 4) The design is described as a randomised control trial, quasi-experimental trial, or pre- and post-intervention i.e. prospective cohort study; 5) The intervention group participated in a structured PA programme, characterised through a recommendation of a defined amount of PA per week achieved through the provision of exercise sessions; 6) Participants were observed for at least six months (in order to be able to consider maintenance as defined above); and 7) The primary aim of the intervention(s) was to increase PA levels. Further criteria include that 8) The intervention(s) needed to

be aimed to change PA behaviour only (single behaviour change), as it was hypothesized that in multi-behaviour change interventions cumulative effects between the intervention components may impact physical activity outcomes; and 9) The intervention(s) did not use mass media and were not described as a home-based or lifestyle intervention, as these strategies would not be easily applicable in a traditional structured setting such as facilities in the private and public health sector.

Data Extraction

Data were extracted via self-designed standardised forms developed by the research team and characteristics of included studies (e.g. type and duration of the intervention(s), follow-up assessments, PA goal or recommendation, applied behaviour change techniques) and study subjects (e.g. gender, age, body mass index (BMI)) were recorded.

Adoption and maintenance rates, and intervention effectiveness expressed as changes in PA levels were defined as primary outcomes. Adoption was defined as the proportion of participants meeting guidelines for PA levels (i.e. 150 minutes of moderate to vigorous activity per week) at the first time point measured after the initiation of the structured PA interventions. Maintenance was considered as the proportion of participants meeting guidelines for PA levels at any time point after the first and exceeding at least six months since the initiation of the structure PA intervention. Intervention effectiveness was examined with respect to the reported difference in the magnitude of effects in influencing PA levels between the structured intervention(s) and the control condition. Three categories of methods were to define and evaluate PA levels were considered: 1) Changes in objectively measured PA; 2) Changes in an intended outcome of PA behaviour as a proxy (i.e. change in VO_{2max}); and 3) Changes in self-reported PA levels.

Within each of the studies the Behaviour Change techniques (BCTs) applied in each intervention were coded. Two members of the research team independently coded the BCTs on the base of the description of interventions published in each article using the taxonomy of Michie et al. (2013) and discrepancies were resolved through discussions[12]. We originally planned to consider sub-group analyses based upon the number of BCTs used in structured PA interventions; however, due to the low numbers of studies

and considerable heterogeneity in designs, and general lack of reporting of proportions in particular, we did not pursue this further. We did still conduct the coding for BCTs in order to aid in the narrative interpretation and synthesis of studies; these are included in the supplementary materials.

Risk of bias

The Cochrane risk of bias tool was used to detect biases in random sequence generation, allocation concealment, blinding of outcome assessors, attrition and reporting[13]. We did not assess bias for blinding study personal and assessors for which intervention a participant received, due to the nature of PA interventions.

For each study, every item was graded into 'high risk' or 'low risk' of bias, or 'unclear' where description of methodology was insufficient. Studies were subjectively graded into high, medium or low risk of bias, considering the types of bias and their severity in addition to the perceived overall methodological complexity. If one item was evaluated as high risk in combination with unclear and low risks for the remaining four categories the overall study quality was assumed to be medium, whereas more than two categories graded as high risk resulted in a low overall quality of the study. Trials for which the risk of random sequence generation has been classified as high were also be given an overall high risk for bias, as this is a potentially strong confounding factor[14]. An overall rating of low risk of bias and therefore high methodological quality was given to studies for which at least three categories were classified as low risk in combination with an unclear risk of bias for the remaining two items. Any study for which the risk assessment of at least three items was not possible due to insufficiently reported methods was be given a moderate overall rating. No studies were excluded due to a poor rating of methodological quality.

Exploratory Meta-analysis

The potential for conducting a meta-analysis was included in the original pre-registration; though, as we anticipated low numbers of studies and considerable heterogeneity in designs, and reporting of proportions in particular, the specific approach was not detailed further. A number of subgroup analyses were also originally considered (Comparison of ARM rates of studies using different (amounts of)

BCTs; ARM rates depending on definitions (and measurements); two types of structured programmes: strict prescription vs. flexible, self-determined approach with recommendation; RCTs vs. quasi-experimental studies; Academic literature (scientific journals) vs. grey literature (evaluation reports); Participant characteristics, e.g. age groups, gender, or ethnicity); however, due to the anticipated issues noted, we did not pursue these further. Instead, meta-analysis was performed in an exploratory manner to examine just the estimates (and their precision) for effectiveness of structured PA interventions upon PA levels during adoption and maintenance, the proportions of those achieving recommended PA levels during adoption and maintenance, the retention of participants at follow-up, and adherence.

For PA levels, where data could be extracted to support between group comparisons (i.e. between structured PA interventions and a control intervention) standardised mean differences (using pooled pre- and post- standard deviations) between groups for either changes in PA levels (post- minus pre-), or PA levels at a given time point, were examined. Some studies only reported within group data for structured PA interventions and for these the standardised mean changes between pre- and post-measures (using pre- standard deviations and an assumed pre- to post-test correlation of 0.7) were examined. An overall standardised mean effect estimate was produced using a multi-level mixed effects model where both research study and intra-study groups were included as random effects. Estimates were weighted by inverse sampling variance to account for the within- and between-study variance (tau-squared). Restricted maximal likelihood estimation was used in all models. This was done for effectiveness during adoption, and maintenance. The magnitude of standardised effects were interpreted with reference to Cohen's thresholds: trivial (<0.2), small (0.2 to <0.5), moderate (0.5 to <0.8), and large (>0.8)[15]. Positive effect size values indicated effects in favour of the structured PA interventions.

For all proportions, where possible data was extracted for structured PA interventions and control interventions. This was considered as the number of participants available for follow-up assessments compared to the number randomised at baseline prior to the intervention initiation. Though for some studies proportions were only reported for structured PA interventions. Thus, two multi-level mixed effects models of the logit transformed proportions where produced with both research study and intra-study groups were included as random effects: one for structured PA interventions and one control

interventions. These two models were then compared and also their estimates were back transformed to raw proportions for reporting. Estimates were weighted by inverse sampling variance to account for the within- and between-study variance (tau-squared). Restricted maximal likelihood estimation was used in all models. This was done for proportions of participants adopting, maintaining, retained, and for adherence.

All analysis was performed using the 'meta' and 'metafor' packages in R (v 3.6.1; R Core Team, https://www.r-project.org/). Point estimates for pooled effect sizes and precision of those estimates using 95% CI are reported.

Results

A total of 20,659 articles were identified through database searching. After duplicates were removed 17,008 articles were checked for eligibility through title and abstract screening, and while for 16,849 articles the title was deemed as not relevant or information in the abstract contradicted the inclusion criteria, 159 articles were analysed through full-text screening (of which two described the same study at different time points[16,17]). After the exclusion of 147 articles which did not meet inclusion criteria the reference lists of the remaining twelve included articles were screened, which yielded the identification of five additional articles. Of those, one was included in this review, resulting in a final inclusion of 13 papers reporting twelve studies[16,17,26,27,18–25]. The screening process is outlined in the flow chart in Figure 1.

(Insert figure 1 here.)

The search of grey literature yielded eight evaluation reports summarising a multitude of PA programmes in the UK. None of those met all of the inclusion criteria, therefore this information was not included in this review.

Study characteristics

The characteristics of the twelve included studies are summarised in Table 1. Studies were published between 1982 and 2017 and most commonly conducted in the USA (50.0%) [16–18,21,25,26,28]. The design

of the studies was predominantly described as a randomised controlled trial (RCT) (41.2%) [16,17,21,22,28,29], and less than half of included articles were of a high methodological quality (n=5, 41.67%) [16–18,21,26,28,29]. Supplementary file 1 provides more information on the risk of bias assessment.

A total of 2116 participants of an average age of 49 years were recruited in the studies included in this review, ranging from 7 to 338 participants in the intervention groups and 7 to 183 in the control groups. For four studies (33.3%) there was no age reported [21,24,27,28]. However, these articles were still included in this review as in the methods of each article it was specified participants had to be of an age of 65 or below in order to be eligible for the respective study. Participants in four studies (33.3%) had a predominantly white background[16,17,25,28,29], whereas two studies only included Latinas (16.6%)[18,21] and one study included African Americans (8.3%)[26]. Half of the studies included women only[21,22,26,27,29], and in the remaining six articles 58.7% of participants were female.

The interventions were of a mean duration of 10±5 months and participants were observed for 14±6 months. In four studies (33.3%) the intervention consisted of a combination of group sessions and self-directed PA[22–24,27], two interventions were held in a church-setting[18,26], continuous group sessions were offered by three research groups (25.0%)[19,21,28,29], and another two studies (16.6%) started in a group setting but were continued as self-directed PA programmes[16,17,25]. Participants were advised to be active ranging from 60 to 270 minutes per week in one to six supervised and unsupervised sessions. Most interventions were compared to a home-based programme (n=3, 25.0%)[16,17,25,29] or no treatment control groups (n=3, 25.0%)[22,23,28], and participants in two control groups were put on a waiting-list to receive the intervention after the data collection was completed[19,27]. In one study, there was no control condition, while the remaining three control groups received either cancer screening[18], safety education[21], or stretching classes[26]. A comprehensive list of all included BCTs for interventions can be found in Supplement 2.

Change in PA levels

To assess effectiveness of structured interventions for increasing PA at the time of adoption, results at the first follow-up assessment were summarised. Three categories of methods used to define and

evaluate PA levels were identified: 1) Changes in objectively measured PA; 2) Changes in an intended outcome of PA behaviour as a proxy (i.e. change in VO_{2max}); and 3) Changes in self-reported PA levels.

Changes in objectively measured PA levels were assessed by only one study using accelerometery. Three studies examined changes in VO_{2max} through maximal treadmill test as a proxy for PA behaviour (i.e. it was inferred that an increase in cardiorespiratory fitness was indicative of a change in PA levels). Different questionnaires were used to determine self-reported PA levels, e.g. 7-day recalls, the Modified Baecke Questionnaire or the Voorips questionnaire.

On average, the first follow-up assessment took place after 5.92 months (range: 2-12). A total of six structured interventions were shown to be more effective than the controls[18,21,23,27–29] (Table 2). Further, Kukkonen et al. (1982) reported significant increases in PA levels in the intervention group[24]. However, these findings were not compared to a control group. Three trials resulted in an increase in PA outcomes in both treatment groups, indicating the effectiveness of both the control condition and the structured intervention[16,17,19,22]. Only one study resulted in an effect in the control group[25], while in another study no effect was found in the intervention or control group[26]. Despite most studies reporting significantly greater effects for the structured PA interventions, during adoption (pre- to first time-point) only a trivial non-significant standardised effect was observed (0.15 [95%CIs = -0.06 to 0.36], k = 12; Figure 2). We did not identify differences in control conditions, PA recommendations, or study quality, between short-term effective and non-effective interventions. All but one[26] of the six interventions selectively including women[18,21,22,27,29] were effective.

To assess maintenance of PA levels, we summarised follow-up assessments from studies occurring at least six months after short-term effects were assessed. Of seven studies, differences in effects were still significant in four trials, in addition to the maintenance of PA levels in the intervention group reported by Kettunen et al. (1982) and both the intervention and control group in the trial conducted by Dunn et al (1997, 1999). This was assessed after an average of 17.88 months (range: 12-24). The overall standardised effect estimate of effectiveness on PA levels during maintenance was also trivial with a wide interval estimate (0.19 [95%CIs = -0.68 to 1.07], k = 7; Figure 3).

(Insert figure 2 and 3 here.)

Adoption and maintenance rates

Information on the proportion of participants reaching the PA recommendations from authors within the studies (which notably varied between studies and often diverged from current guidelines; see table 1) was only provided for three of twelve studies.

Adoption and maintenance behaviour were assessed via the same methods as for physical activity levels, with exemption of two studies using activity recalls for the analysis of PA levels but assessed the stages of change for the evaluation of adoption and maintenance rates. The Stages of Change are also referred to as the 'Transtheoretical Model' and describe a theory aiming to explain behaviour change or somebody's readiness for change. It is based on the assumption each individual moves through five stages when aiming for sustainably change his or her behaviour, namely precontemplation, contemplation, preparation, action, maintenance. Whereas stage 1 and 2 are characterised by someone's intentions to become more active (stage 2) or the absence of such (stage 1), individuals in the preparation stage will undertake first steps towards fulfilling their goal. Stage 4 and 5 describe active individuals who either have (stage 5) or have not (stage 4) been fulfilling their activity goals for at least six months[33]. Therefore, a change from stage 1, 2 or 3 to stage 4 and from stage 4 to 5 describe adoption or maintenance of PA, respectively.

For structured interventions, the three studies that assessed for adoption reported rates of 53%, 63%, or 85% for the intervention group, while for the control group these rates were lower with 16%, 40% and 78% (Table 2). In two cases, this was assessed after six months and one study assessed adoption at twelve months. Meta-analytic estimates of proportions adopting did not differ between structure PA or controls, though estimates had low precision (66.6% [95%CIs = 39.8% to 85.7%], k = 3 vs 42.3% [95%CIs = 12.5% to 78.8%], k = 2 respectively).

Of those three studies, long-term maintenance effects were evaluated in only two with time points of data collection differing largely (12 vs. 24 months). Maintenance rates were reported as 20% and 38% in the intervention group and 20% and 15% in the control conditions. Meta-analytic estimates of

proportions maintaining also did not differ between structure PA or controls, though the estimate for structured PA interventions in particular had low precision (28.4% [95%CIs = 13.8% to 49.5%], *k* = 2 vs 18.3% [95%CIs = 13.5% to 24.3%], *k* = 2 respectively).

Retention Rates and Dropout

Retention rates were published in ten articles (Table 3). For eight studies providing information on reasons for dropout, no differences were found between control conditions and interventions. Six studies reported the frequency of dropout reasons for 136 participants in the control groups and 156 participants in the intervention groups [16–19,22,25,29]. For the control conditions, most frequently reported reasons for the termination of the participation in the interventions were: Loss to follow-up (n=41); Lack of time and/or motivation (n=54), and withdrawal, illnesses and/or injuries (n=14), and non-interest in the study (n=11). Participants in the intervention groups most commonly dropped out due to a lack of time and/or motivation (n=69); loss to follow-up (n=27); withdrawal (n=20); illnesses and/or injuries (n=14); and other unspecified reasons (n=14). For both the intervention and control groups rarely reported reasons included: relocation (n=3); death (of family member) (n=2); unreliable responses (n=1); pregnancy (n=1). Other reasons where no frequencies were reported were non-participation, lack of spousal support, domestic violence, and missing consent forms[21,26]. Retention at follow-up did not differ between structured PA or controls (75.1% [95%CIs = 65.0% to 83.0%], *k* = 17 vs 75.4% [95%CIs = 67.0% to 82.3%], *k* = 15 respectively).

Adherence Rates

Adherence was assessed by evaluating the proportion of attended intervention sessions in relation to the recommendation (Table 3). This was documented through self-reported exercise logs and registers at the exercise sessions. In five articles adherence to the structured programme was reported [19,21,25,28,29]. In three cases, control conditions providing alternative sessions were assessed for adherence. Adherence to the structured programmes was similar between included studies, with the exemption of Young and Steward (2006), who reported an adherence rate of 18%[26]. For the control conditions Yang et al. (2016) report an adherence rate differing majorly from the average, as participants adhered to 100% to the activity recommendation, in contrast to the intervention group with 67%[27]. The overall estimates for

adherence also did not differ between structured PA and controls (63.0% [95%CIs = 55.6% to 69.6%], k = 5 vs 77.8% [95%CIs = 19.4% to 98.1%], k = 2 respectively).

Discussion

This review highlights limited evidence in support of the effectiveness of structured PA interventions to improve PA levels compared to control interventions (including home-based interventions, screening, education, stretching, or non-intervention controls including wait-lists). In addition, it suggests that adherence to structured PA interventions may be similar when compare to control interventions that are not merely non-intervention or wait-lists. Further, it demonstrates the general lack of reporting regarding adoption and maintenance rates of structured PA interventions limiting interpretation of intervention effectiveness and feasibility.

The discrepancy between the assessment of intervention feasibility by considering proportions of individuals adopting and maintaining behaviours, alongside effectiveness, is highlighted when interpreting the results published by Dunn et al. (1999), who showed that despite a maintenance rate of only 20%, a significant increase in PA levels at follow-up compared to baseline levels was observed[16,17]. Dunn et al. (1999) applied different questionnaires to assess individual and group effects, and implemented a study that hasn't been advanced by many other researchers in over two decades. The use of different questionnaires might partly explain the variance of findings, though this further emphasises the need for consistency in reporting. This example does however serve to underline how group level changes may be found to be statistically significant even when only a small proportion of individuals in the sample groups have considerable improvements. From a public health perspective, though increased population average PA levels are no doubt desirable, most would want for this to occur as a result of large proportions of the population meeting recommendations. Thus, merely considering whether a statistically significant change, even a change of large magnitude, in a chosen outcome measure such as PA levels occurs may not reflect the relevance and magnitude of effects and ultimately the interventions effectiveness. That being said, our exploratory meta-analytic effect estimates suggested that any effects of structured PA interventions upon PA levels is likely trivial to small at best; though there was significant study effect heterogeneity (Q(df = 11) = 30.0109, p = 0.0016).

The mechanisms of why some PA interventions are effective, whereas others do not result in the anticipated effect are not fully understood. Many studies individually found a tendency for structured programmes to result in a greater increase in PA levels than the control conditions in short-term, with more varying results in long-term. Another recent meta-analysis found PA interventions lasting between 9-15 months to result in higher PA levels than controls (standardised mean difference (SMD) = 0.20; 95% CI (0.13-0.26))., providing evidence for their effectiveness in long-term[30]. Howlett et al (2019) recently published a meta-analysis including both interventions aiming to increase activity levels and reduce sedentary time in healthy adults, and suggested that PA interventions were effective in short- (SMD = 0.32, 95% CI (0.16-0.48)), and long-term (SMD = 0.21, 95% CI (0.12-0.30))[31]. They further conducted meta-regressions, analysing the associations of BCTs influencing intervention effectiveness. A BCT is described as an 'active ingredient' of an intervention, detailing how a targeted behaviour is intended to be changed[18]. Interventions usually consist of a combination of different BCTs and the analysis of patterns in effective or non-effective interventions can therefore contribute to understanding the mechanisms of each intervention[32]. As such, consideration of the BCTs included within interventions may aid in the understanding of which are most effective for enhancing adoption, retention, and maintenance.

In our review we originally intended to consider analysis of adoption, maintenance, retention, and adherence based upon the application of BCTs; however, the heterogeneity of studies in addition to the general lack of reporting of such outcomes led us to not conduct this. Considering the general lack of effectiveness of structured PA interventions upon PA levels, combined with the heterogeneity of individual intervention effects, to draw clearer conclusions of the effectiveness of specific BCTs applied in structured interventions to increase PA levels, more empirical research comparing specific approaches to facilitate meta-analysis is warranted.

More information and ultimately more research are needed to systematically summarise adoption and maintenance rates of PA interventions. Of twelve interventions only three were assessed for adoption and or maintenance rates. These three individually reported the interventions to be successful in inducing adoption in 50-85% in participants. The length of time to the first follow-up measurement

varied widely though and indeed some studies could be considered to be capturing the long-term effects at their first assessment. The absence of a prior assessment though renders it impossible to know whether this reflects the 'maintenance' of a behaviour after adoption has occurred as the latter must be first observed to infer the former. Where long-term maintenance of behaviours was assessed after adoption and the termination of the programmes, only 20-35% of participants were able to sustain effects and were therefore classified as maintainers. Further, though the small number of studies renders the effect estimates imprecise, we could not identify clear differences in adoption or maintenance between structured PA interventions or controls within our meta-analysis. The generally low maintenance rates observed in this review indicate the need for future research to improve the effectiveness of interventions in order to increase the proportion of participants meeting the desired behaviours.

The effectiveness of an intervention should further be interpreted in light of a participant retention and adherence rate. In this review, we found that retention at follow-up was no better for structured PA interventions compared to controls (75.1% vs 75.4% respectively); equating to an approximate dropout rate of ~25%. In a meta-analysis of yoga interventions an average dropout rate of 11% was observed, rising to 15% for interventions of a duration of twelve weeks or longer[38]. Similar results have been published by Stubbs et al. (2016) in their meta-analysis, showing that 18% of participants with depression do not complete the full course of a PA programme[33]. This lies in the recommended range of up to 20% dropout, as specified by Cochrane guidelines[13]. Although our results show a slightly higher dropout, our findings are in accordance with the dropout rates found in PA interventions in clinical populations e.g. schizophrenia (27%) and HIV (29%) patients[34,35]. However, as we only included interventions recruiting healthy individuals, the comparability to these other reviews is limited.

Adherence for structured PA interventions compared to control interventions requiring session attendance did not differ (63.0% vs 77.8% respectively; though few studies reported attendance for control conditions resulting in imprecision for this estimate. Adherence of course impacts upon the fidelity of the interventions, and the heterogeneity observed in regard to the effectiveness of structured interventions is perhaps likely to be caused by both different intervention designs in addition to non-

compliance of participants. Compared to the other included studies, Young and Steward (2006), reported an adherence rate marginally lower than the average[18]. In fact, this trial was the only one selectively including African American women, and one of two trials implementing a culturally adapted church-based programme. As the recommendation of being active for 60 minutes per week lies at the lower end of the spectrum of recommendations in the included studies, this is unlikely to have resulted in the low adherence rate. Historically, African Americans as a demographic are reported to be less active than white adults[36,37]. Common barriers to PA among African American women are lack of motivation, family obligations and lack of social support, and haircare maintenance and the preference of a more voluminous body shape, which are less commonly found in any other ethnic group[38]. This indicates African American Women may require additional support to become and stay physically active, and might explain why adherence in this study was lower compared to the other included articles.

For the control conditions Yang et al. (2016) report an adherence rate differing majorly from the average, as participants adhered to 100% to the activity recommendation, in contrast to the intervention group with 67%[27]. This suggests the same amount of PA was more easily integrated into an individual's weekly routine in a home-based programme alone than a group-based programme that transitioned into a home-based programme after two months, as described for the intervention group in this trial. However, this was considered a pilot trial, and therefore this hypothesis warrants further examination.

This review has several limitations. Firstly, we selectively included structured programmes, as they are most commonly applied with respect to PA interventions and, due to their standardisation, are more readily replicated[3]. However, while searching for literature, the identification of an intervention as structured was often difficult due to insufficient reporting of methodological design and thus there may be studies that went unidentified and were thus excluded. This insufficient reporting also impaired the risk of bias assessment and the coding for BCTs, although we perceive there to have been an improvement in the descriptions of more recent publications. For each included treatment group, it is possible more BCTs have been applied than we coded for, due to insufficient descriptions of intervention design. Future studies should follow appropriate taxonomy for describing the inclusion of

BCTs in interventions, in addition to following the Consensus on Exercise Reporting Template[3] when reporting on structured physical activity/exercise interventions.

Our ability to draw clear conclusions from this review is impaired by the heterogeneity of structured PA interventions, highlighting the second limitation of this review. Recommended PA levels, the content and delivery of exercise sessions, assessment of PA levels, control conditions, intervention duration and observation period are amongst the multitude of characteristics of studies differing considerably between trials, resulting in a large number of factors potentially influencing intervention effectiveness. Combined with the relative lack of studies, our quantitative synthesis using meta-analysis therefore did not permit meta-regression and subgroup analyses taking intervention characteristics into account. Thus, we are unable to explore the mechanisms associated with effective structured interventions.

We further highlight the issue of using surrogate measures of PA like maximal treadmill tests for the assessment of effectiveness of PA interventions. Those measures only provide indirect insights on PA levels where it is assumed that, where a PA intervention is delivered compared to a control, any improvements in cardiorespiratory fitness will only be due to increased PA levels and thus can be used as a surrogate marker of this behaviour. While this might appear reasonable and indeed improvements in cardiorespiratory fitness are an often and intended outcome of performance as a result of PA behaviour being linked to morbidity and mortality[39], many factors might influence changes in cardiorespiratory fitness including genetics[40] as well as both the volume and intensity of effort of any PA behaviours[41,42]. As such it is argued for future work that PA behaviours, and indeed the fidelity of any intervention with respect to the PA behaviours (volume, intensity of effort, frequency etc.) should be assessed in addition to the intended outcome of those behaviours (e.g. cardiorespiratory fitness, strength, mental wellbeing etc.). This would permit greater understanding of both what impact interventions have upon PA behaviours and subsequently the degree to which those behaviours might mediate improved health, fitness, and wellbeing.

We originally intended to compare academic literature with grey literature in the form of evaluation reports of structured public health interventions. However, we were unable to identify non-academic literature outlining PA interventions in sufficient detail to be included in this review. We believe evaluation of public health interventions can contribute to the improvement of current knowledge on effective PA interventions, however, lack of control conditions and poor reporting has historically been an issue within the sector (PP 1), though does seem to have improved (PP2) and thus this may be a possibility in future review[42].

We originally perceived our strict inclusion criteria as a strength rather than a limitation, as this was intended to enable us to draw focused conclusions on the effectiveness of structured interventions by reducing heterogeneity of PA programmes. However, despite this we still found considerable heterogeneity in interventions. We suggest for future reviews to include a broader spectrum of interventions to be able to carry out a more comprehensive review and meta-analysis, and to conduct sub-analyses and meta-regressions where appropriate. For example, this could include interventions such as brief advice which is often recommended as a cost-effective approach[43]. A further strength of this review is our adherence to PRISMA and AMSTAR 2 guidelines to produce a review to the recommended standard. Thirdly, we pre-registered the research protocol for this review on PROSPERO prior to conducting the search, enabling researchers conducting similar reviews to understand, reproduce or improve our approach. Moreover, our results can be used to inform future evidence-based structured interventions.

Conclusion

From the limited data available it can be concluded that structured PA interventions at best have a trivial to small effect upon PA levels. Thus, structured PA interventions appear to lack evidence for effectiveness in improving PA levels. Further, though retention is often reported and is similar between interventions and controls, adoption, maintenance and adherence rates were rarely reported rendering difficulty in interpreting results of effectiveness of structured PA interventions.

To our knowledge this is the first review aiming to systematically summarise adoption, retention maintenance, and adherence rates of long-term structured PA interventions in non-clinical adults. As these rates are rarely reported, we propose a new point of view in regards of the evaluation of studies considering these, and highly recommend future research to address this issue of underreporting by publishing information on adoption and maintenance rates relative to the recommended amount of PA. This will contribute to the improvement of our understanding of the feasibility and implementation of PA interventions, the mechanism through which they are effective in changing PA behaviour, and therefore the design of future PA interventions aiming to tackle global inactivity rates. A more comprehensive summary and meta-analysis of future literature is needed, including a wider range of PA interventions.

Acknowledgements

None.

Funding Sources

This project was funded by the ukactive Research Institute and Coventry University.

References

1. World Health Organisation (WHO). Prevalence of insufficient physical activity among adults Data by WHO region. Published 2018. https://apps.who.int/gho/data/view.main.2482?lang=en

2. Sport England. Active Lives Adult Survey - November 17/18 report. Published 2019. https://www.sportengland.org/media/13898/active-lives-adult-november-17-18-report.pdf

3. Slade SC, Dionne CE, Underwood M, et al. Consensus on Exercise Reporting Template (CERT): Modified Delphi Study. *Phys Ther*. 2016;96(10):1514-1524. doi:10.2522/ptj.20150668 [doi]

4. World Health Organisation (WHO). Noncommunicable diseases and mental health; Target 3: Reduce prevalence of physical inactivity. Published 2018. http://www.who.int/nmh/ncd-tools/target3/en/

5. Kahlert D. Maintenance of physical activity: Do we know what we are talking about? *Prev Med Rep*. 2015;2:178-180. doi:10.1016/j.pmedr.2015.02.013

6. Marcus BH, Dubbert PM, Forsyth LH, et al. Physical Activity Behavior Change: Issues in Adoption and Maintenance. Krantz DS, Wing RR, Voorhees C, Hill DR, eds. *Heal Psychol*. 2000;19(1):32-41. doi:10.1037/0278-6133.19.Suppl1.32

7. Fjeldsoe B, Neuhaus M, Winkler E, Eakin E. Systematic review of maintenance of behaviour change following physical activity and dietary interventions. *Health Psychol*. 2011;30(1):99-109. doi:10.1037/a0021974 [doi]

8. Prochaska JO, DiClemente CC. Transtheoretical therapy: toward a more integrative model of change. *Psychother Theory, Res Pract*. 1982;19(3):276-288.

9. Howlett N, Trivedi D, Troop NA, Chater AM. Are physical activity interventions for healthy inactive adults effective in promoting behavior change and maintenance, and which behavior change techniques are effective? A systematic review and meta-analysis. *Transl Behav Med*. 2019;9(1):147-157.

10. Liberati A, Altman DG, Tetzlaff J, et al. The PRISMA statement for reporting systematic reviews and meta-analyses of studies that evaluate healthcare interventions: explanation and elaboration. *BMJ*. 2009;339(b2700).

11. Shea BJ, Reeves BC, Wells G, et al. AMSTAR 2: a critical appraisal tool for systematic reviews that include randomised or non-randomised studies of healthcare interventions, or both. *BMJ*. 2017;358(j4008).

12. Michie S, Richardson M, Johnston M, et al. The behavior change technique taxonomy (v1) of 93 hierarchically clustered techniques: building an international consensus for the reporting of behavior change interventions. *Ann Behav Med*. 2013;46(1):81-95. doi:10.1007/s12160-013-9486-6 [doi]

13. Furlan AD, Pennick V, Bombardier C, van Tulder M, Editorial Board CBRG. 2009 updated method guidelines for systematic reviews in the Cochrane Back Review Group. *Spine (Phila Pa 1976)*. 2009;34(18):1929-1941. doi:10.1097/BRS.0b013e3181b1c99f [doi]

14. Higgins JPT, Green S. *Cochrane Handbook for Systematic Reviews of Interventions Version 5.1.0 [Updated March 2011]*. The Cochrane Collaboration; 2011. http://www.handbook.cochrane.org

15. Cohen J. *Statistical Power Analysis for the Behavioral Sciences*. 2nd ed. Routledge; 1988. doi:10.4324/9780203771587

16. Dunn AL, Marcus BH, Kampert JB, Garcia ME, Kohl 3rd HW, Blair SN. Reduction in

cardiovascular disease risk factors: 6-month results from Project Active. *Prev Med (Baltim)*. 1997;26(6):883-892. doi:S0091-7435(97)90218-8 [pii]

17. Dunn AL, Marcus BH, Kampert JB, Garcia ME, Kohl HW, Blair SN. Comparison of Lifestyle and structured interventions to increase physical activity and cardiorespiratory fitness. *JAMA*. 1999;281(4):327-334.

18. Young DR, Stewart KJ. A church-based physical activity intervention for African American women. *Fam Community Heal J Heal Promot Maint*. 2006;29(2):103-117. doi:10.1097/00003727-200604000-00006

19. Lee C, White SW. Controlled trial of a minimal-intervention exercise program for middle-aged working women. *Psychol Health*. 1997;12(3):361-374. doi:10.1080/08870449708406713

20. Arredondo EM, Elder JP, Haughton J, et al. Fe en Accion: Promoting Physical Activity Among Churchgoing Latinas. *Am J Public Health*. 2017;107(7):1109-1115. doi:10.2105/AJPH.2017.303785 [doi]

21. De Jong J, Lemmink KAPM, Stevens M, et al. Six- month effects of the Groningen active living model (GALM) on physical activity, health and fitness outcomes in sedentary and underactive older adults aged 55– 65. *Patient Educ Couns*. 2006;62(1):132-141. doi:10.1016/j.pec.2005.06.017

22. Cox TJ, Puddey TJ, Burke TJ, Beilin TJ, Gorely TJ. Exercise behaviour change in 40 to 65-year- old women: The SWEAT Study (Sedentary Women Exercise Adherence Trial). *Br J Health Psychol*. 2003;8(4):477-495. doi:10.1348/135910703770238329

23. Hovell MF, Mulvihill MM, Buono MJ, et al. Culturally Tailored Aerobic Exercise Intervention for Low-income Latinas. *Am J Heal Promot*. 2008;22(3):155-163. http://search.ebscohost.com/login.aspx?direct=true&db=sph&AN=29999953&site=ehost-live

24. Hertogh EM, Vergouwe Y, Schuit AJ, Peeters PHM, Monninkhof EM. Behavioral Changes after a 1-yr Exercise Program and Predictors of Maintenance. *Med Sci Sport Exerc*. 2010;42(5). http://journals.lww.com/acsm-msse/Fulltext/2010/05000/Behavioral_Changes_after_a_1_yr_Exercise_Program.7.aspx

25. KETTUNEN O, VUORIMAA T, VASANKARI T. A 12-month exercise intervention decreased stress symptoms and increased mental resources among working adults - Results perceived after a 12-month follow-up. *Int J Occup Med Environ Heal*. 2015;28(1):157-168. doi:10.13075/ijomeh.1896.00263

26. Kukkonen K, Rauramaa R, Siitonen O, Hänninen O. Physical training of obese middle-aged persons. *Ann Clin Res*. 1982;14(34):80-85.

27. Yang K, James KA. Yoga, as a transitional platform to more active lifestyle: A 6-month pilot study in the USA. *Health Promot Int*. 2016;31(2):423-429. doi:10.1093/heapro/dau108

28. King A, Haskell W, Young D, Oka R, Stefanick M. Long-term effects of varying intensities and formats of physical activity on participation rates, fitness, and lipoproteins in men and women aged 50 to 65 years. *Circulation*. 1995;91(10):2596.

29. Cox KL, Burke V, Gorely TJ, Beilin LJ, Puddey IB. Controlled comparison of retention and adherence in home- vs center-initiated exercise interventions in women ages 40-65 years: the S.W.E.A.T. Study (Sedentary Women Exercise Adherence Trial). *Prev Med (Baltim)*. 2003;36(1):17-29. http://search.ebscohost.com/login.aspx?direct=true&db=ccm&AN=106837321&site=ehost-live

30. Murray JM, Brennan SF, French DP, Patterson CC, Kee F, Hunter RF. Effectiveness of physical activity interventions in achieving behaviour change maintenance in young and

middle aged adults: A systematic review and meta-analysis. *Soc Sci Med*. 2017;192:125-133. doi:S0277-9536(17)30556-7 [pii]

31. Howlett N, Trivedi D, Troop NA, Chater AM. Are physical activity interventions for healthy inactive adults effective in promoting behavior change and maintenance, and which behavior change techniques are effective? A systematic review and meta-analysis. *Transl Behav Med*. 2019;9(1):147-157. doi:10.1093/tbm/iby010 [doi]

32. Michie S, Abraham C, Eccles MP, Francis JJ, Hardeman W, Johnston M. Strengthening evaluation and implementation by specifying components of behaviour change interventions: a study protocol. *Implement Sci*. 2011;6(10).

33. Stubbs B, Vancampfort D, Rosenbaum S, et al. Dropout from exercise randomized controlled trials among people with depression: A meta-analysis and meta regression. *J Affect Disord*. 2016;190:457-466. doi:S0165-0327(15)30882-X [pii]

34. Vancampfort D, Mugisha J, Richards J, et al. Dropout from physical activity interventions in people living with HIV: a systematic review and meta-analysis. *AIDS Care*. 2017;29(5):636-643. doi:10.1080/09540121.2016.1248347 [doi]

35. Vancampfort D, Rosenbaum S, Schuch FB, Ward PB, Probst M, Stubbs B. Prevalence and predictors of treatment dropout from physical activity interventions in schizophrenia: a meta-analysis. *Gen Hosp Psychiatry*. 2016;39:15-23. doi:10.1016/j.genhosppsych.2015.11.008 [doi]

36. Joseph RP, Ainsworth BE, Keller C, Dodgson JE. Barriers to Physical Activity Among African American Women: An Integrative Review of the Literature. *Women Health*. 2015;55(6):679-699. doi:10.1080/03630242.2015.1039184 [doi]

37. American Heart Association. Heart Disease and Stroke Statistics—2006 Update. *Circulation*. 2006;113(6):e85-e151.

38. Ross R, Blair SN, Arena R, et al. Importance of Assessing Cardiorespiratory Fitness in Clinical Practice: A Case for Fitness as a Clinical Vital Sign: A Scientific Statement From the American Heart Association. *Circulation*. 2016;134(24):e653-e699. doi:CIR.0000000000000461 [pii]

39. Peter I, Papandonatos GD, Belalcazar LM, et al. Genetic modifiers of cardiorespiratory fitness response to lifestyle intervention. *Med Sci Sports Exerc*. 2014;46(2):302-311. doi:10.1249/MSS.0b013e3182a66155 [doi]

40. Montero D, Lundby C. Refuting the myth of non-response to exercise training: "non-responders" do respond to higher dose of training. *J Physiol*. 2017;595(11):3377-3387. doi:10.1113/JP273480 [doi]

41. Ross R, de Lannoy L, Stotz PJ. Separate Effects of Intensity and Amount of Exercise on Interindividual Cardiorespiratory Fitness Response. *Mayo Clin Proc*. 2015;90(11):1506-1514. doi:10.1016/j.mayocp.2015.07.024 [doi]

42. Public Health England. Identifying what works for local physical inactivity interventions. Published 2014. http://researchinstitute.ukactive.com/downloads/managed/Identifying_what_works.pdf

43. NICE (National Institute for Health and Care Excellence). *Physical Activity: Brief Advice for Adults in Primary Care*.; 2013. Accessed June 19, 2020. https://www.nice.org.uk/guidance/ph44/resources/physical-activity-brief-advice-for-adults-in-primary-care-pdf-1996357939909

Tables

Author (Year)	Country	Type of Trial	Study Quality	Sample size	Age (Years)	% Female	Ethnicity	Setting	Intervention Duration	Recommendation	Control
Arredondo et al. (2017)	USA	Cluster-RCT	high	I: 187 C: 183	I:44.5±9.8 C:44.4±9.4	100	Latino	Culturally adapted church-based intervention	12 months	150 minutes MVPA, individual choice of the amount of sessions (Walking groups, cardio dance and strength training classes)	Cancer screening classes
Cox et al. (2003)	Australia	RCT	high	I: 64 C: 62	I: 48.0 C: 48.3	100	Pre-dominantly white	Group-based programme	18 months	3 times or more per week 20 minutes or longer MVPA (walks, aerobics, circuits)	Home-based
De Jong et al. (2006)	Nether-lands	Cluster-RCT	moderate	I: 79 C: 102	I: 59.6±2.4 C: 58.8±2.7	I: 54.4 C: 56.9	NIL	Gym-based leisure-time programme,	6 months	One session of 60 min per week, (the 15 most popular sports for older adults)	Waiting list
Dunn et al. (1997; 1999)	USA	RCT	high	I: 121 C: 114	I: 46.2±6.5 C: 45.9±6.8	I: 50.9 C: 50	Pre-dominantly white	Group-based programme, (after 6 months home-based)	6 months (24 months observation)	150 min MVPA, stage 4 or 5 of the TTM initially 3 sessions, then 5 session per week, duration of 20-60 min, 50-85% of maximal aerobic power	Home-based
Hertogh et al. (2010)	Nether-lands	RCT	moderate	I: 78 C: 68	I: 59.0±4.5 C: 58.7±4.0	100	NIL	Group-based and individual sessions	12 months (24 months observation)	3 sessions per week, twice supervised (60 min, strength and endurance), once individual (30 min brisk walking or cycling)	No intervention
Hovell et al. (2008)	USA	RCT	high	I: 68 C: 66	NIL	100	Latino	Group-based programme for low-income, monolingual Spanish speaking Latinas	6 months (12 months observation)	3 times 90 minutes per week (aerobic dance exercise)	Safety education

Kettunen et al. (2015)	Finland	C-ITS	low	I: 338 C: 33	I: 45±8.8 C: 41±6.9	I: 62.7 C: 51.5	NIL	Supervised and individual exercise sessions' (after 12 months unsupervised exercise)	12 months (24 months observation)	2-day training camps at assessments (1-2 supervised sessions per 5 months), 3-5 unsupervised sessions per week (walking, skiing, biking)	No intervention
King et al. (1995)	USA	RCT	high	I: 74 C: 75	NIL	I: 45.9 C: 45.3	Pre-dominantly white	Group-based intervention	12 months	3 sessions of 60 min per week (walking/jogging, treadmills and stationary cycles)	No intervention
Kukkonen et al. (1982)	Finland	Cohort	Low	I: 169	NILz	57.4	NIL	supervised and individual sessions, individual training programme for each participant	17 months	30-60 min 3-6 times a week (walking, skiing, jogging, swimming, cycling), one a week supervised (calisthenics, volleyball)	No control group
Lee et al. (1997)	Australia	Cross-over	moderate	I: 14 C: 11	NIL	100	NIL	self-administered programme of low-impact aerobic exercise, weekly activity and education	3 months (12 months observation)*	weekly low intensity aerobic exercise session, (duration unknown, walking or exercise to music) 2-3 individual sessions	Waiting list
Yang et al. (2016)	USA	Pilot RCT	low	I: 7 C: 7	I: 58.4±6.8 C: 58.7±4.1	I: 85.7 C: 85.7	Pre-dominantly NIL (missing information). white	pilot study, group-based programme (after 2 months home-based)	6 months	1 supervised session (90 min) and 2 unsupervised sessions at least twice a week, yoga	Home-based (Yoga DVD), 2 initial supervised sessions
Young and Steward (2006)	USA	Cluster RCT	high	I: 123 C: 73	I:48.2±24.4 C:48.4±19.7	100	African American	Culturally adapted church-based aerobic intervention	6 months	individualised activity plans, 60 min per week	alternating weekly low-intensity stretching classes and health lectures

Table 1: Study Design and Intervention Characteristics. *Due to the cross-over design of this study only data until 24 weeks were included in this review. I (Intervention), C (Control).

	Adoption				Maintenance			
Author (Year)	**PA Measure**	**Time point [m]**	**Rate [%]**	**Short-term increase in PA levels**	**PA Measure**	**Time point [m]**	**Rate [%]**	**Long-term increase in PA levels**
Objectively measured physical activity								
Arredondo et al. (2017)	Accelerometer	12	I: 45 C: 34	Intervention group		No follow-up		
Physical activity behaviour outcome as proxy								
Kettunen et al. (2015) [B]	Maximal Treadmill test (VO_{2max})	4	NIL	Intervention group	Maximal Treadmill test (VO_{2max})	12, 24	NIL	Intervention group (both assessments)
King et al. (1995) [A, B]	Maximal Treadmill test (VO_{2max})	6	NIL	Intervention group	Maximal Treadmill test (VO_{2max})	12	NIL	Intervention group
Kukkonen et al. (1982) [B]	Maximal Treadmill test (VO_{2max})	2	NIL	Intervention group*	Maximal Treadmill test (VO_{2max})	17	NIL	Intervention group*
Stages of Change								
Cox et al. (2003) [C]	7-day Recall (energy expenditure)	6	unclear	Intervention group	Stage of Change instrument (stage 5)	18	unclear	No difference
Dunn et al. (1997; 1999) [+]	7-day recall (6m), Maximal treadmill test (24m)	6	I: 85 C: 77	Both groups	Proportion of participants meeting the ACSM recommendation (stage 5)	24	I: 20 C: 20	Both groups
Changes in self-reported physical activity behaviour								
De Jong et al. (2006)	Voorrips questionnaire (energy expenditure)	6	NIL	Both groups		No follow-up		
Hertogh et al. (2010)	Modified Baecke Questionnaire (MET/h)	12	NIL	Both groups	Modified Baecke Questionnaire (MET/h)	24	NIL	Intervention group
Hovell et al. (2008)	7-day recall	6	I: 63 C: 17	Intervention group	7-day recall	12	I: 38 C: 15	Intervention group
Lee et al. (1997)	Physical activity recall (MVPA)	3	NIL	Intervention group		Not assessed		
Yang et al. (2016)	Modifiable Activity Questionnaire (MET/h)	2	NIL	Control group		No follow-up		
Young and Steward (2006)	Physical activity recall	6	NIL	None		No follow-up		

Table 2: Adoption and Maintenance of Physical Activity. [A]Only information of two groups included in this review. [B]Data from two publications summarised. * No control group. Physical activity (PA); Intervention (I); Control (C); Moderate to vigorous physical activity (MVPA); Metabolic equivalents of tasks per hour (MET/h); NIL (missing information).

Author (Year)	Retention Rate			Overall Adherence Rate	
	Follow-up (months)	Intervention	Control	Intervention	Control
Arredondo et al. (2017)	12	87.56%	86.76%	NIL	NIL
Cox et al. (2003)	18	81.25%	61.29%	65.38%	50.85%
	12	90.63%	66.13%		
	6	96.88%	90.32%		
De Jong et al. (2006)	6	48.47%	67.11%	80.00%	NIL
Dunn et al. (1997; 1999) [A]	24	78.26 %	81.97%	NIL	NIL
	12	92.98%	95.87%		
Hertogh et al. (2010)	24	81.25%	73.12%	NIL	NIL
Hovell et al. (2008)	12	NIL	NIL	64.10%	50.00%
Kettunen et al. (2015)	24	84.85%	52.66%	NIL	NIL
	12	90.53%	84.85%		
	8	90.53%	81.82%		
	4	81.66%	87.88%		
King et al. (1995) [A]	12	NIL	NIL	52.56 %	NIL
Kukkonen et al. (1982)	17	(men): 56.94% (women): 55.67%	NIL	NIL	NIL
Lee et al. (1997)	12	73.68%	72.22%	NIL	NIL
Yang et al. (2016)	6	85.71%	57.14%	67.31%	100%
Young and Steward (2006)	6	60.98%	57.53%	NIL	NIL

Table 3: Retention and Adherence Rates. [A]Only information of two groups included in this review. NIL (missing information).

Figures

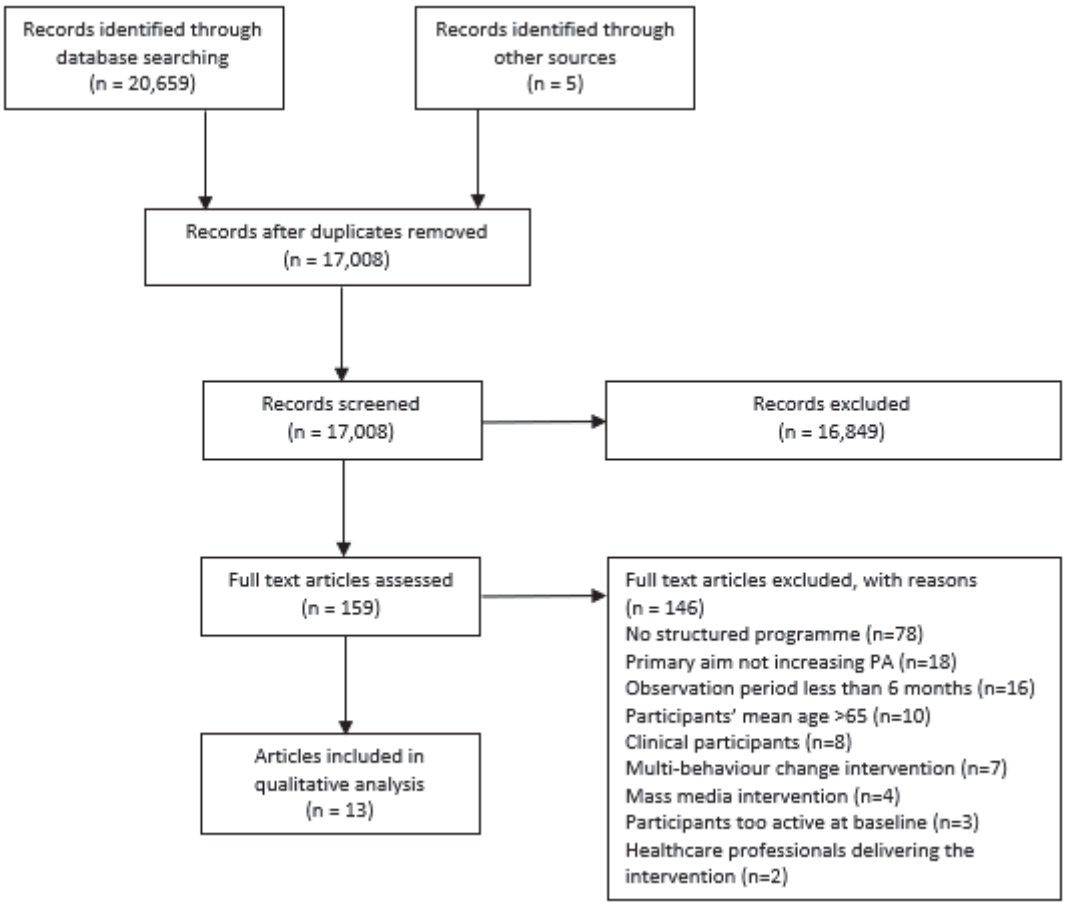

Figure 1: PRISMA Chart Screening Process

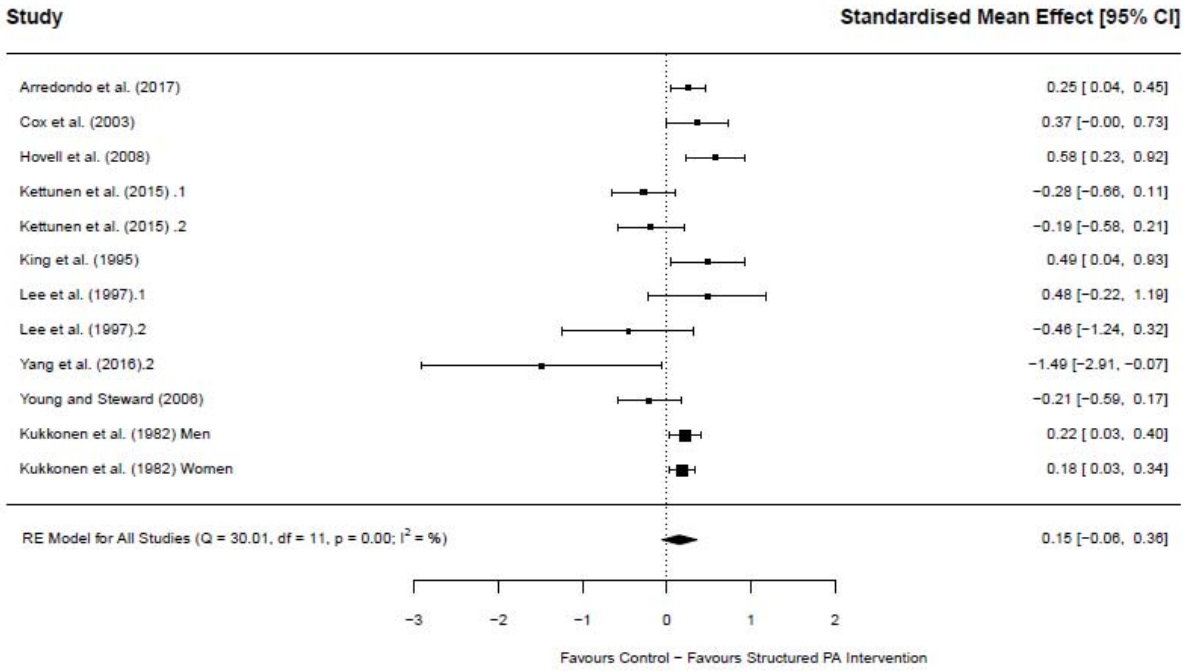

Figure 2: Forest Plot of the standardised mean effect of structured interventions compared to controls at adoption.

The study conducted by Kettunen et al. (2015) model included two assessments investigating adoption (1: 4 months, 2: 8 months). This also accounts for Lee et al. (1997) (1:2 months, 2: 6 months), and Yang et al. (2016) (1:2 months, 2:4 months).

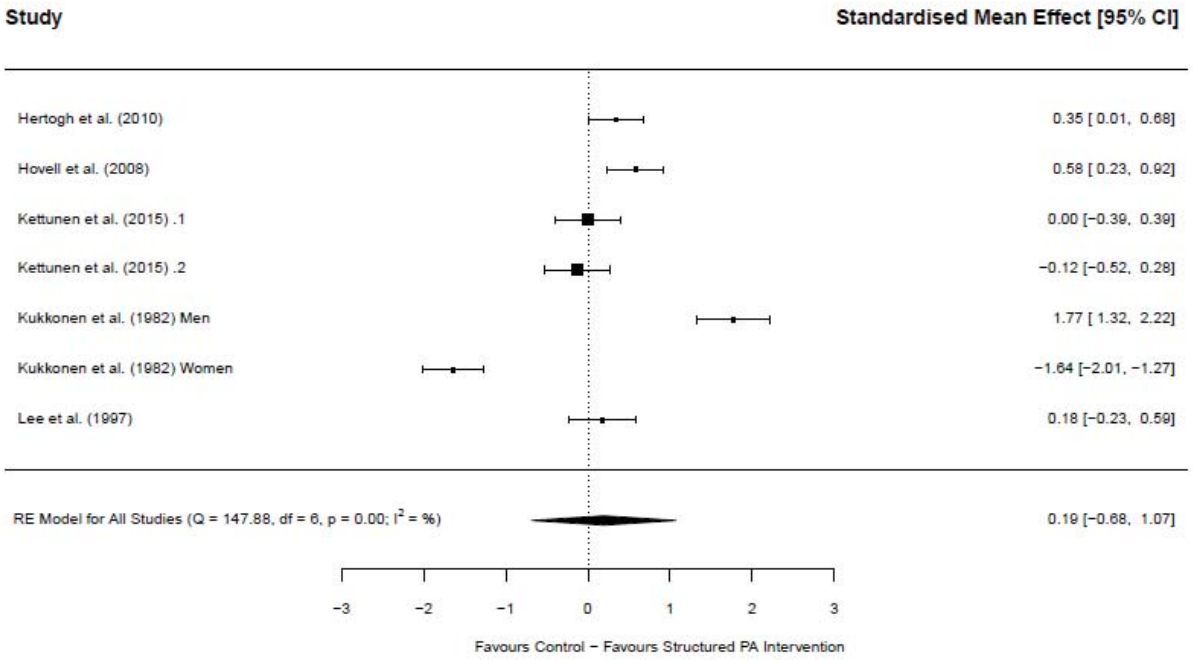

Figure 3: Forest Plot of the standardised mean effect of structured interventions compared to controls six months after the assessment of adoption (i.e. maintenance).

The study conducted by Kettunen et al. (2015) included two assessments investigating maintenance (1: 12 months, 2: 24 months).